Patent Screen

Your Fasta Input (protein sequence):

>CL001305, SEQ ID NO:2, March 3, 2004

```
    1 MAFARRLLRG PLSGPLLGRR GVCAGAMAPP CRFVLELPDC TLAHFALGAD
   51 APGDADAPDP RLAALLGPPE RSYSLCVPVT PDAGCGARVR AARLHQRLLH
  101 QLRRGPFQRC QLLRLLCYCP GGQAGGAQQG FLLRDPLDDP DTRQALLELL
  151 GACQEAPRPH LGEFEADPRG QLWQRLWEVQ DGRRLQVGCA QVVPVPEPPL
  201 HPVVPDLPSS VVFPDREAAR AVLEECTSFI PEARAVLDLV DQCPKQIQKG
  251 KFQVVAIEGL DATGKTTVTQ SVADSLKAVL LKSPPSCIGQ WRKIFDDEPT
  301 IIRRAFYSLG NYIVASEIAK ESAKSPVIVD RYWHSTATYA IATEVSGGLQ
  351 HLPPAHHPVY QWPEDLLKPD LILLLTVSPE ERLQRLQGRG MEKTREEAEL
  401 EANSVFRQKV EMSYQRMENP GCHVVDASPS REKVLQTVLS LIQNSFSEP
```


FastAlertP Blast Report:

BLASTP 2.2.5 [Nov-16-2002]
Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer,
Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997),
"Gapped BLAST and PSI-BLAST: a new generation of protein database search
programs", Nucleic Acids Res. 25:3389-3402.

Query= CL001305, SEQ ID NO:2, March 3, 2004
        (449 letters)

Database: FastAlert_P.fasta
          221,820 sequences; 37,667,424 total letters

Sequences producing significant alignments:	Score (bits)	E Value
CRA|222000011464779 /dataset=FastAlert_P /length=449 /altid=Derw...	909	0.0
CRA|224000047885450 /dataset=FastAlert_P /length=422 /altid=Derw...	842	0.0
CRA|224000047885452 /dataset=FastAlert_P /length=251 /altid=Derw...	455	e-128
CRA|179000322886965 /dataset=FastAlert_P /length=200 /altid=Derw...	382	e-106
CRA|179000322886965 /dataset=FastAlert_P /length=200 /altid=Derw...	382	e-106

```
>CRA|222000011464779 /dataset=FastAlert_P /length=449
          /altid=Derwent|WO2004009778.17 /def=/PT="New human
          kinases and phosphatases, useful for diagnosing,
          treating or preventing atherosclerosis, hypertension,
          AIDS, allergy, multiple sclerosis, osteoarthritis,
          Alzheimer's disease, Crohn's disease, cancer or
          hepatitis" /QU="DPT" /PA="INCYTE CORP" /PD="29-JAN-2004"
          /PR="19-JUL-2002" /ED="19-FEB-2004"
        Length = 449
```

 Score = 909 bits (2350), Expect = 0.0
 Identities = 448/449 (99%), Positives = 448/449 (99%)

```
Query: 1    MAFARRLLRGPLSGPLLGRRGVCAGAMAPPCRFVLELPDCTLAHFALGADAPGDADAPDP 60
            MAFARRLLRGPLSGPLLGRRGVCAGAMAPP RFVLELPDCTLAHFALGADAPGDADAPDP
Sbjct: 1    MAFARRLLRGPLSGPLLGRRGVCAGAMAPPRRFVLELPDCTLAHFALGADAPGDADAPDP 60

Query: 61   RLAALLGPPERSYSLCVPVTPDAGCGARVRAARLHQRLLHQLRRGPFQRCQLLRLLCYCP 120
            RLAALLGPPERSYSLCVPVTPDAGCGARVRAARLHQRLLHQLRRGPFQRCQLLRLLCYCP
Sbjct: 61   RLAALLGPPERSYSLCVPVTPDAGCGARVRAARLHQRLLHQLRRGPFQRCQLLRLLCYCP 120

Query: 121  GGQAGGAQQGFLLRDPLDDPDTRQALLELLGACQEAPRPHLGEFEADPRGQLWQRLWEVQ 180
```


1

```
            GGQAGGAQQGFLLRDPLDDPDTRQALLELLGACQEAPRPHLGEFEADPRGQLWQRLWEVQ
Sbjct: 121  GGQAGGAQQGFLLRDPLDDPDTRQALLELLGACQEAPRPHLGEFEADPRGQLWQRLWEVQ 180

Query: 181  DGRRLQVGCAQVVPVPEPPLHPVVPDLPSSVVFPDREAARAVLEECTSFIPEARAVLDLV 240
            DGRRLQVGCAQVVPVPEPPLHPVVPDLPSSVVFPDREAARAVLEECTSFIPEARAVLDLV
Sbjct: 181  DGRRLQVGCAQVVPVPEPPLHPVVPDLPSSVVFPDREAARAVLEECTSFIPEARAVLDLV 240

Query: 241  DQCPKQIQKGKFQVVAIEGLDATGKTTVTQSVADSLKAVLLKSPPSCIGQWRKIFDDEPT 300
            DQCPKQIQKGKFQVVAIEGLDATGKTTVTQSVADSLKAVLLKSPPSCIGQWRKIFDDEPT
Sbjct: 241  DQCPKQIQKGKFQVVAIEGLDATGKTTVTQSVADSLKAVLLKSPPSCIGQWRKIFDDEPT 300

Query: 301  IIRRAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVY 360
            IIRRAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVY
Sbjct: 301  IIRRAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVY 360

Query: 361  QWPEDLLKPDLILLLTVSPEERLQRLQGRGMEKTREEAELEANSVFRQKVEMSYQRMENP 420
            QWPEDLLKPDLILLLTVSPEERLQRLQGRGMEKTREEAELEANSVFRQKVEMSYQRMENP
Sbjct: 361  QWPEDLLKPDLILLLTVSPEERLQRLQGRGMEKTREEAELEANSVFRQKVEMSYQRMENP 420

Query: 421  GCHVVDASPSREKVLQTVLSLIQNSFSEP 449
            GCHVVDASPSREKVLQTVLSLIQNSFSEP
Sbjct: 421  GCHVVDASPSREKVLQTVLSLIQNSFSEP 449


>CRA|224000047885450 /dataset=FastAlert_P /length=422
            /altid=Derwent|WO2003102159.1112 /def=/PT="New NOVX
            polypeptides and nucleic acid molecules useful for
            preventing or treating NOVX-associated disorders, e.g.
            cancer, cardiomyopathy, atherosclerosis or diabetes, in
            chromosome mapping, tissue typing or in
            pharmacogenomics" /QU="DPT" /PA="CURAGEN CORP"
            /PD="11-DEC-2003" /PR="04-JUN-2002" /ED="25-DEC-2003"
            Length = 422

 Score =  842 bits (2174), Expect = 0.0
 Identities = 420/424 (99%), Positives = 420/424 (99%), Gaps = 3/424 (0%)

Query: 27   MAPPCRFVLELPDCTLAHFALGADAPGDADAPDPRLAALLGPPERSYSLCVPVTPDAGCG 86
            MAPP RFVLELPDCTLAHFALGADAPGDADAPDPRLAALLGPPERSYSLCVPVTPDAGCG
Sbjct: 1    MAPPRRFVLELPDCTLAHFALGADAPGDADAPDPRLAALLGPPERSYSLCVPVTPDAGCG 60

Query: 87   ARVRAARLHQRLLHQLRRGPFQRCQLLRLLCYCPGGQAGGAQQGFLLRDPLDDPDTRQAL 146
            ARVRAARLHQRLLHQLRRGPFQRCQLLRLLCYCPGGQAGGAQQGFLLRDPLDDPDTRQAL
Sbjct: 61   ARVRAARLHQRLLHQLRRGPFQRCQLLRLLCYCPGGQAGGAQQGFLLRDPLDDPDTRQAL 120

Query: 147  LELLGACQEAPRPHLGEFEADPRGQLWQRLWEVQDGRRLQVGCAQVVPVPEPPLHPVVPD 206
            LELLGACQEAPRPHLGEFEADPRGQLWQRLWEVQDGRRLQVGCAQVVPVPEPPLHPVVPD
Sbjct: 121  LELLGACQEAPRPHLGEFEADPRGQLWQRLWEVQDGRRLQVGCAQVVPVPEPPLHPVVPD 180

Query: 207  LPSSVVFPDREAARAVLEECTSFIPEARAVLDLVDQCPKQIQKGKFQVVAIEGLDAT-GK 265
            LPSSVVFPDREAARAVLEECTSFIPEARAVLDLVDQCPKQIQKGKFQVVAIEGLDAT GK
Sbjct: 181  LPSSVVFPDREAARAVLEECTSFIPEARAVLDLVDQCPKQIQKGKFQVVAIEGLDATGGK 240

Query: 266  TTVTQSVADSLKAVLLKSPPSCIGQWRKIFDDEPTIIRRAFYSLGNYIVASEIAKESAKS 325
            TTVTQSVADSLKAVLLKSPPSCIGQWRKIFDDEPTIIRRAFYSLGNYIVASEIAKESAKS
Sbjct: 241  TTVTQSVADSLKAVLLKSPPSCIGQWRKIFDDEPTIIRRAFYSLGNYIVASEIAKESAKS 300

Query: 326  PVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVYQWPEDLLKPDLILLLTVSPEERLQR 385
            PVIVDR  HSTATYAIATEVSGGLQHLPPAHHPVYQWPEDLLKPDLILLLTVSPEERLQR
Sbjct: 301  PVIVDR--HSTATYAIATEVSGGLQHLPPAHHPVYQWPEDLLKPDLILLLTVSPEERLQR 358

Query: 386  LQGRGMEKTREEAELEANSVFRQKVEMSYQRMENPGCHVVDASPSREKVLQTVLSLIQNS 445
            LQGRGMEKTREEAELEANSVFRQKVEMSYQRMENPGCHVVDASPSREKVLQTVLSLIQNS
```

2

```
Sbjct: 359  LQGRGMEKTREEAELEANSVFRQKVEMSYQRMENPGCHVVDASPSREKVLQTVLSLIQNS 418

Query: 446  FSEP 449
            FSEP
Sbjct: 419  FSEP 422
```

>CRA|224000047885452 /dataset=FastAlert_P /length=251
        /altid=Derwent|WO2003102159.1118 /def=/PT="New NOVX
        polypeptides and nucleic acid molecules useful for
        preventing or treating NOVX-associated disorders, e.g.
        cancer, cardiomyopathy, atherosclerosis or diabetes, in
        chromosome mapping, tissue typing or in
        pharmacogenomics" /QU="DPT" /PA="CURAGEN CORP"
        /PD="11-DEC-2003" /PR="04-JUN-2002" /ED="25-DEC-2003"
        Length = 251

 Score = 455 bits (1171), Expect = e-128
 Identities = 229/232 (98%), Positives = 229/232 (98%)

```
Query: 218  AARAVLEECTSFIPEARAVLDLVDQCPKQIQKGKFQVVAIEGLDATGKTTVTQSVADSLK 277
            A  AVLEECTSFIPEARAVLDLVDQCPKQIQKGKFQVVAIEGLDATGKTTVTQS ADSLK
Sbjct: 20   ALGAVLEECTSFIPEARAVLDLVDQCPKQIQKGKFQVVAIEGLDATGKTTVTQSAADSLK 79

Query: 278  AVLLKSPPSCIGQWRKIFDDEPTIIRRAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTA 337
            AVLLKSPPSCIGQWRKIFDDEPTIIRRAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTA
Sbjct: 80   AVLLKSPPSCIGQWRKIFDDEPTIIRRAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTA 139

Query: 338  TYAIATEVSGGLQHLPPAHHPVYQWPEDLLKPDLILLLTVSPEERLQRLQGRGMEKTREE 397
            TYAIATEVSGGLQHLPPAHHPVYQWPEDLLKPDLILLLTVSPEERLQRLQGRGMEKTREE
Sbjct: 140  TYAIATEVSGGLQHLPPAHHPVYQWPEDLLKPDLILLLTVSPEERLQRLQGRGMEKTREE 199

Query: 398  AELEANSVFRQKVEMSYQRMENPGCHVVDASPSREKVLQTVLSLIQNSFSEP 449
            AELEANSVFRQKVEMSYQRMENPGCHVVDASPSREKVLQTVLSLIQNSFSEP
Sbjct: 200  AELEANSVFRQKVEMSYQRMENPGCHVVDASPSREKVLQTVLSLIQNSFSEP 251
```

>CRA|179000322886965 /dataset=FastAlert_P /length=200
        /altid=Derwent|WO2003102159.1116 /def=/PT="New NOVX
        polypeptides and nucleic acid molecules useful for
        preventing or treating NOVX-associated disorders, e.g.
        cancer, cardiomyopathy, atherosclerosis or diabetes, in
        chromosome mapping, tissue typing or in
        pharmacogenomics" /QU="DPT" /PA="CURAGEN CORP"
        /PD="11-DEC-2003" /PR="04-JUN-2002" /ED="25-DEC-2003"
        Length = 200

 Score = 382 bits (981), Expect = e-106
 Identities = 193/195 (98%), Positives = 194/195 (99%), Gaps = 1/195 (0%)

```
Query: 256  AIEGLDATG-KTTVTQSVADSLKAVLLKSPPSCIGQWRKIFDDEPTIIRRAFYSLGNYIV 314
            +IEGLDATG KTTVTQSVADSLKAVLLKSPPSCIGQWRKIFDDEPTIIRRAFYSLGNYIV
Sbjct: 3    SIEGLDATGGKTTVTQSVADSLKAVLLKSPPSCIGQWRKIFDDEPTIIRRAFYSLGNYIV 62

Query: 315  ASEIAKESAKSPVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVYQWPEDLLKPDLILL 374
            ASEIAKESAKSPVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVYQWPEDLLKPDLILL
Sbjct: 63   ASEIAKESAKSPVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVYQWPEDLLKPDLILL 122

Query: 375  LTVSPEERLQRLQGRGMEKTREEAELEANSVFRQKVEMSYQRMENPGCHVVDASPSREKV 434
            LTVSPEERLQRLQGRGMEKTREEAELEANSVFRQKVEMSYQRMENPGCHVVDASPSREKV
Sbjct: 123  LTVSPEERLQRLQGRGMEKTREEAELEANSVFRQKVEMSYQRMENPGCHVVDASPSREKV 182

Query: 435  LQTVLSLIQNSFSEP 449
```

```
                 LQTVLSLIQNSFSEP
Sbjct: 183 LQTVLSLIQNSFSEP 197


>CRA|179000322886965 /dataset=FastAlert_P /length=200
           /altid=Derwent|WO2003102159.1114 /def=/PT="New NOVX
           polypeptides and nucleic acid molecules useful for
           preventing or treating NOVX-associated disorders, e.g.
           cancer, cardiomyopathy, atherosclerosis or diabetes, in
           chromosome mapping, tissue typing or in
           pharmacogenomics" /QU="DPT" /PA="CURAGEN CORP"
           /PD="11-DEC-2003" /PR="04-JUN-2002" /ED="25-DEC-2003"
           Length = 200

 Score =  382 bits (981), Expect = e-106
 Identities = 193/195 (98%), Positives = 194/195 (99%), Gaps = 1/195 (0%)

Query: 256 AIEGLDATG-KTTVTQSVADSLKAVLLKSPPSCIGQWRKIFDDEPTIIRRAFYSLGNYIV 314
           +IEGLDATG KTTVTQSVADSLKAVLLKSPPSCIGQWRKIFDDEPTIIRRAFYSLGNYIV
Sbjct: 3   SIEGLDATGGKTTVTQSVADSLKAVLLKSPPSCIGQWRKIFDDEPTIIRRAFYSLGNYIV 62

Query: 315 ASEIAKESAKSPVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVYQWPEDLLKPDLILL 374
           ASEIAKESAKSPVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVYQWPEDLLKPDLILL
Sbjct: 63  ASEIAKESAKSPVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVYQWPEDLLKPDLILL 122

Query: 375 LTVSPEERLQRLQGRGMEKTREEAELEANSVFRQKVEMSYQRMENPGCHVVDASPSREKV 434
           LTVSPEERLQRLQGRGMEKTREEAELEANSVFRQKVEMSYQRMENPGCHVVDASPSREKV
Sbjct: 123 LTVSPEERLQRLQGRGMEKTREEAELEANSVFRQKVEMSYQRMENPGCHVVDASPSREKV 182

Query: 435 LQTVLSLIQNSFSEP 449
           LQTVLSLIQNSFSEP
Sbjct: 183 LQTVLSLIQNSFSEP 197


  Database: FastAlert_P.fasta
    Posted date:  Mar 2, 2004  5:57 AM
  Number of letters in database: 37,667,424
  Number of sequences in database:  221,820

Lambda     K       H
   0.321    0.138    0.426

Gapped
Lambda     K       H
   0.267   0.0410    0.140


Matrix: BLOSUM62
Gap Penalties: Existence: 11, Extension: 1
Number of Hits to DB: 39,731,498
Number of Sequences: 221820
Number of extensions: 1943285
Number of successful extensions: 4818
Number of sequences better than 1.0e-08: 5
Number of HSP's better than  0.0 without gapping: 5
Number of HSP's successfully gapped in prelim test: 0
Number of HSP's that attempted gapping in prelim test: 4808
Number of HSP's gapped (non-prelim): 6
length of query: 449
length of database: 37,667,424
effective HSP length: 106
effective length of query: 343
effective length of database: 14,154,504
```

effective search space: 4854994872
effective search space used: 4854994872
T: 11
A: 40
X1: 16 (7.4 bits)
X2: 38 (14.6 bits)
X3: 64 (24.7 bits)
S1: 41 (21.8 bits)
S2: 141 (58.9 bits)

GENESEQP Blast Report:

BLASTP 2.2.5 [Nov-16-2002]
Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer,
Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997),
"Gapped BLAST and PSI-BLAST: a new generation of protein database search
programs", Nucleic Acids Res. 25:3389-3402.

Query= CL001305, SEQ ID NO:2, March 3, 2004
        (449 letters)

Database: geneseqp_all.fasta
          1,546,707 sequences; 273,213,384 total letters

```
                                                             Score     E
Sequences producing significant alignments:                 (bits)  Value
CRA|224000046444906 /dataset=GENESEQP /org=Homo sapiens /taxon=9...   914   0.0
CRA|224000046444906 /dataset=GENESEQP /org=Homo sapiens /taxon=9...   914   0.0
CRA|222000011464779 /dataset=GENESEQP /org=Homo sapiens /taxon=9...   909   0.0
CRA|224000047885450 /dataset=GENESEQP /org=Homo sapiens /taxon=9...   842   0.0
CRA|98000043611837 /dataset=GENESEQP /org=Mus musculus /taxon=10...   602   e-171
CRA|224000047885452 /dataset=GENESEQP /org=Homo sapiens /taxon=9...   455   e-127
CRA|223000060714134 /dataset=GENESEQP /org=Homo sapiens /taxon=9...   441   e-123
CRA|114000021936825 /dataset=GENESEQP /org=Mus musculus /taxon=1...   259   2e-68
CRA|224000046444908 /dataset=GENESEQP /org=Mus musculus /taxon=1...   199   3e-50
```

>CRA|224000046444906 /dataset=GENESEQP /org=Homo sapiens /taxon=9606
          /mol_type=aa /date=18-DEC-03 /length=449
          /altid=derwent_ac|ADC63370 /altid=derwent_id|ADC63370
          /def=Human novel kinase protein /patent=US2003087294-A1
          /pat_section=Claim
        Length = 449

 Score = 914 bits (2362), Expect = 0.0
 Identities = 449/449 (100%), Positives = 449/449 (100%)

```
Query: 1    MAFARRLLRGPLSGPLLGRRGVCAGAMAPPCRFVLELPDCTLAHFALGADAPGDADAPDP 60
            MAFARRLLRGPLSGPLLGRRGVCAGAMAPPCRFVLELPDCTLAHFALGADAPGDADAPDP
Sbjct: 1    MAFARRLLRGPLSGPLLGRRGVCAGAMAPPCRFVLELPDCTLAHFALGADAPGDADAPDP 60

Query: 61   RLAALLGPPERSYSLCVPVTPDAGCGARVRAARLHQRLLHQLRRGPFQRCQLLRLLCYCP 120
            RLAALLGPPERSYSLCVPVTPDAGCGARVRAARLHQRLLHQLRRGPFQRCQLLRLLCYCP
Sbjct: 61   RLAALLGPPERSYSLCVPVTPDAGCGARVRAARLHQRLLHQLRRGPFQRCQLLRLLCYCP 120

Query: 121  GGQAGGAQQGFLLRDPLDDPDTRQALLELLGACQEAPRPHLGEFEADPRGQLWQRLWEVQ 180
            GGQAGGAQQGFLLRDPLDDPDTRQALLELLGACQEAPRPHLGEFEADPRGQLWQRLWEVQ
Sbjct: 121  GGQAGGAQQGFLLRDPLDDPDTRQALLELLGACQEAPRPHLGEFEADPRGQLWQRLWEVQ 180

Query: 181  DGRRLQVGCAQVVPVPEPPLHPVVPDLPSSVVFPDREAARAVLEECTSFIPEARAVLDLV 240
            DGRRLQVGCAQVVPVPEPPLHPVVPDLPSSVVFPDREAARAVLEECTSFIPEARAVLDLV
Sbjct: 181  DGRRLQVGCAQVVPVPEPPLHPVVPDLPSSVVFPDREAARAVLEECTSFIPEARAVLDLV 240
```

```
Query:  241  DQCPKQIQKGKFQVVAIEGLDATGKTTVTQSVADSLKAVLLKSPPSCIGQWRKIFDDEPT  300
             DQCPKQIQKGKFQVVAIEGLDATGKTTVTQSVADSLKAVLLKSPPSCIGQWRKIFDDEPT
Sbjct:  241  DQCPKQIQKGKFQVVAIEGLDATGKTTVTQSVADSLKAVLLKSPPSCIGQWRKIFDDEPT  300

Query:  301  IIRRAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVY  360
             IIRRAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVY
Sbjct:  301  IIRRAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVY  360

Query:  361  QWPEDLLKPDLILLLTVSPEERLQRLQGRGMEKTREEAELEANSVFRQKVEMSYQRMENP  420
             QWPEDLLKPDLILLLTVSPEERLQRLQGRGMEKTREEAELEANSVFRQKVEMSYQRMENP
Sbjct:  361  QWPEDLLKPDLILLLTVSPEERLQRLQGRGMEKTREEAELEANSVFRQKVEMSYQRMENP  420

Query:  421  GCHVVDASPSREKVLQTVLSLIQNSFSEP  449
             GCHVVDASPSREKVLQTVLSLIQNSFSEP
Sbjct:  421  GCHVVDASPSREKVLQTVLSLIQNSFSEP  449


>CRA|224000046444906 /dataset=GENESEQP /org=Homo sapiens /taxon=9606
          /mol_type=aa /date=10-JUN-03 /length=449
          /altid=derwent_ac|ABU08678 /altid=derwent_id|ABU08678
          /def=Novel human thymidylate kinase subfamily kinase
          /patent=US6489153-B1 /pat_section=Claim
          Length = 449

 Score =  914 bits (2362), Expect = 0.0
 Identities = 449/449 (100%), Positives = 449/449 (100%)

Query:  1    MAFARRLLRGPLSGPLLGRRGVCAGAMAPPCRFVLELPDCTLAHFALGADAPGDADAPDP  60
             MAFARRLLRGPLSGPLLGRRGVCAGAMAPPCRFVLELPDCTLAHFALGADAPGDADAPDP
Sbjct:  1    MAFARRLLRGPLSGPLLGRRGVCAGAMAPPCRFVLELPDCTLAHFALGADAPGDADAPDP  60

Query:  61   RLAALLGPPERSYSLCVPVTPDAGCGARVRAARLHQRLLHQLRRGPFQRCQLLRLLCYCP  120
             RLAALLGPPERSYSLCVPVTPDAGCGARVRAARLHQRLLHQLRRGPFQRCQLLRLLCYCP
Sbjct:  61   RLAALLGPPERSYSLCVPVTPDAGCGARVRAARLHQRLLHQLRRGPFQRCQLLRLLCYCP  120

Query:  121  GGQAGGAQQGFLLRDPLDDPDTRQALLELLGACQEAPRPHLGEFEADPRGQLWQRLWEVQ  180
             GGQAGGAQQGFLLRDPLDDPDTRQALLELLGACQEAPRPHLGEFEADPRGQLWQRLWEVQ
Sbjct:  121  GGQAGGAQQGFLLRDPLDDPDTRQALLELLGACQEAPRPHLGEFEADPRGQLWQRLWEVQ  180

Query:  181  DGRRLQVGCAQVVPVPEPPLHPVVPDLPSSVVFPDREAARAVLEECTSFIPEARAVLDLV  240
             DGRRLQVGCAQVVPVPEPPLHPVVPDLPSSVVFPDREAARAVLEECTSFIPEARAVLDLV
Sbjct:  181  DGRRLQVGCAQVVPVPEPPLHPVVPDLPSSVVFPDREAARAVLEECTSFIPEARAVLDLV  240

Query:  241  DQCPKQIQKGKFQVVAIEGLDATGKTTVTQSVADSLKAVLLKSPPSCIGQWRKIFDDEPT  300
             DQCPKQIQKGKFQVVAIEGLDATGKTTVTQSVADSLKAVLLKSPPSCIGQWRKIFDDEPT
Sbjct:  241  DQCPKQIQKGKFQVVAIEGLDATGKTTVTQSVADSLKAVLLKSPPSCIGQWRKIFDDEPT  300

Query:  301  IIRRAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVY  360
             IIRRAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVY
Sbjct:  301  IIRRAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVY  360

Query:  361  QWPEDLLKPDLILLLTVSPEERLQRLQGRGMEKTREEAELEANSVFRQKVEMSYQRMENP  420
             QWPEDLLKPDLILLLTVSPEERLQRLQGRGMEKTREEAELEANSVFRQKVEMSYQRMENP
Sbjct:  361  QWPEDLLKPDLILLLTVSPEERLQRLQGRGMEKTREEAELEANSVFRQKVEMSYQRMENP  420

Query:  421  GCHVVDASPSREKVLQTVLSLIQNSFSEP  449
             GCHVVDASPSREKVLQTVLSLIQNSFSEP
Sbjct:  421  GCHVVDASPSREKVLQTVLSLIQNSFSEP  449


>CRA|222000011464779 /dataset=GENESEQP /org=Homo sapiens /taxon=9606
          /mol_type=aa /date=04-DEC-03 /length=449
          /altid=derwent_ac|AAE38613 /altid=derwent_id|AAE38613
```

```
                /def=Human thymidylate kinase protein
                /patent=WO2003064642-A1 /pat_section=Claim
                Length = 449


 Score =  909 bits (2350), Expect = 0.0
 Identities = 448/449 (99%), Positives = 448/449 (99%)


Query: 1    MAFARRLLRGPLSGPLLGRRGVCAGAMAPPCRFVLELPDCTLAHFALGADAPGDADAPDP 60
            MAFARRLLRGPLSGPLLGRRGVCAGAMAPP RFVLELPDCTLAHFALGADAPGDADAPDP
Sbjct: 1    MAFARRLLRGPLSGPLLGRRGVCAGAMAPPRRFVLELPDCTLAHFALGADAPGDADAPDP 60


Query: 61   RLAALLGPPERSYSLCVPVTPDAGCGARVRAARLHQRLLHQLRRGPFQRCQLLRLLCYCP 120
            RLAALLGPPERSYSLCVPVTPDAGCGARVRAARLHQRLLHQLRRGPFQRCQLLRLLCYCP
Sbjct: 61   RLAALLGPPERSYSLCVPVTPDAGCGARVRAARLHQRLLHQLRRGPFQRCQLLRLLCYCP 120


Query: 121  GGQAGGAQQGFLLRDPLDDPDTRQALLELLGACQEAPRPHLGEFEADPRGQLWQRLWEVQ 180
            GGQAGGAQQGFLLRDPLDDPDTRQALLELLGACQEAPRPHLGEFEADPRGQLWQRLWEVQ
Sbjct: 121  GGQAGGAQQGFLLRDPLDDPDTRQALLELLGACQEAPRPHLGEFEADPRGQLWQRLWEVQ 180


Query: 181  DGRRLQVGCAQVVPVPEPPLHPVVPDLPSSVVFPDREAARAVLEECTSFIPEARAVLDLV 240
            DGRRLQVGCAQVVPVPEPPLHPVVPDLPSSVVFPDREAARAVLEECTSFIPEARAVLDLV
Sbjct: 181  DGRRLQVGCAQVVPVPEPPLHPVVPDLPSSVVFPDREAARAVLEECTSFIPEARAVLDLV 240


Query: 241  DQCPKQIQKGKFQVVAIEGLDATGKTTVTQSVADSLKAVLLKSPPSCIGQWRKIFDDEPT 300
            DQCPKQIQKGKFQVVAIEGLDATGKTTVTQSVADSLKAVLLKSPPSCIGQWRKIFDDEPT
Sbjct: 241  DQCPKQIQKGKFQVVAIEGLDATGKTTVTQSVADSLKAVLLKSPPSCIGQWRKIFDDEPT 300


Query: 301  IIRRAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVY 360
            IIRRAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVY
Sbjct: 301  IIRRAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVY 360


Query: 361  QWPEDLLKPDLILLLTVSPEERLQRLQGRGMEKTREEAELEANSVFRQKVEMSYQRMENP 420
            QWPEDLLKPDLILLLTVSPEERLQRLQGRGMEKTREEAELEANSVFRQKVEMSYQRMENP
Sbjct: 361  QWPEDLLKPDLILLLTVSPEERLQRLQGRGMEKTREEAELEANSVFRQKVEMSYQRMENP 420


Query: 421  GCHVVDASPSREKVLQTVLSLIQNSFSEP 449
            GCHVVDASPSREKVLQTVLSLIQNSFSEP
Sbjct: 421  GCHVVDASPSREKVLQTVLSLIQNSFSEP 449



>CRA|224000047885450 /dataset=GENESEQP /org=Homo sapiens /taxon=9606
            /mol_type=aa /date=18-DEC-03 /length=422
            /altid=derwent_ac|ADC39136 /altid=derwent_id|ADC39136
            /def=Novel human NOVX polypeptide  SEQ ID NO: 78
            /patent=WO2003010327-A2 /pat_section=Claim
            Length = 422


 Score = 842 bits (2174), Expect = 0.0
 Identities = 420/424 (99%), Positives = 420/424 (99%), Gaps = 3/424 (0%)


Query: 27   MAPPCRFVLELPDCTLAHFALGADAPGDADAPDPRLAALLGPPERSYSLCVPVTPDAGCG 86
            MAPP RFVLELPDCTLAHFALGADAPGDADAPDPRLAALLGPPERSYSLCVPVTPDAGCG
Sbjct: 1    MAPPRRFVLELPDCTLAHFALGADAPGDADAPDPRLAALLGPPERSYSLCVPVTPDAGCG 60


Query: 87   ARVRAARLHQRLLHQLRRGPFQRCQLLRLLCYCPGGQAGGAQQGFLLRDPLDDPDTRQAL 146
            ARVRAARLHQRLLHQLRRGPFQRCQLLRLLCYCPGGQAGGAQQGFLLRDPLDDPDTRQAL
Sbjct: 61   ARVRAARLHQRLLHQLRRGPFQRCQLLRLLCYCPGGQAGGAQQGFLLRDPLDDPDTRQAL 120


Query: 147  LELLGACQEAPRPHLGEFEADPRGQLWQRLWEVQDGRRLQVGCAQVVPVPEPPLHPVVPD 206
            LELLGACQEAPRPHLGEFEADPRGQLWQRLWEVQDGRRLQVGCAQVVPVPEPPLHPVVPD
Sbjct: 121  LELLGACQEAPRPHLGEFEADPRGQLWQRLWEVQDGRRLQVGCAQVVPVPEPPLHPVVPD 180


Query: 207  LPSSVVFPDREAARAVLEECTSFIPEARAVLDLVDQCPKQIQKGKFQVVAIEGLDAT-GK 265
```

7

```
        LPSSVVFPDREAARAVLEECTSFIPEARAVLDLVDQCPKQIQKGKFQVVAIEGLDAT GK
Sbjct: 181 LPSSVVFPDREAARAVLEECTSFIPEARAVLDLVDQCPKQIQKGKFQVVAIEGLDATGGK 240

Query: 266 TTVTQSVADSLKAVLLKSPPSCIGQWRKIFDDEPTIIRRAFYSLGNYIVASEIAKESAKS 325
        TTVTQSVADSLKAVLLKSPPSCIGQWRKIFDDEPTIIRRAFYSLGNYIVASEIAKESAKS
Sbjct: 241 TTVTQSVADSLKAVLLKSPPSCIGQWRKIFDDEPTIIRRAFYSLGNYIVASEIAKESAKS 300

Query: 326 PVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVYQWPEDLLKPDLILLLTVSPEERLQR 385
        PVIVDR  HSTATYAIATEVSGGLQHLPPAHHPVYQWPEDLLKPDLILLLTVSPEERLQR
Sbjct: 301 PVIVDR--HSTATYAIATEVSGGLQHLPPAHHPVYQWPEDLLKPDLILLLTVSPEERLQR 358

Query: 386 LQGRGMEKTREEAELEANSVFRQKVEMSYQRMENPGCHVVDASPSREKVLQTVLSLIQNS 445
        LQGRGMEKTREEAELEANSVFRQKVEMSYQRMENPGCHVVDASPSREKVLQTVLSLIQNS
Sbjct: 359 LQGRGMEKTREEAELEANSVFRQKVEMSYQRMENPGCHVVDASPSREKVLQTVLSLIQNS 418

Query: 446 FSEP 449
        FSEP
Sbjct: 419 FSEP 422


>CRA|98000043611837 /dataset=GENESEQP /org=Mus musculus /taxon=10090
        /mol_type=aa /date=04-DEC-03 /length=395
        /altid=derwent_ac|AAE38614 /altid=derwent_id|AAE38614
        /def=Mouse thymidylate kinase protein
        /patent=WO2003064642-A1 /pat_section=Disclosure;
        Length = 395

 Score =  602 bits (1552), Expect = e-171
 Identities = 294/378 (77%), Positives = 330/378 (87%)

Query: 68  PPERSYSLCVPVTPDAGCGARVRAARLHQRLLHQLRRGPFQRCQLLRLLCYCPGGQAGGA 127
        P   +  LCVP+ P  GCG RV+AAR+H RLL QLRRGP QRCQL +LL Y PG QAG A
Sbjct: 16  PQGAATRLCVPLAPGEGCGPRVQAARVHHRLLQQLRRGPLQRCQLSKLLGYGPGDQAGEA 75

Query: 128 QQGFLLRDPLDDPDTRQALLELLGACQEAPRPHLGEFEADPRGQLWQRLWEVQDGRRLQV 187
        Q GFLLRDP D PDTR+ LL+LLG+CQEA RP L EF+AD +G LWQRLWE+Q  R++QV
Sbjct: 76  QHGFLLRDPCDHPDTRRDLLQLLGSCQEAARPQLAEFQADSQGLLWQRLWELQGDRQVQV 135

Query: 188 GCAQVVPVPEPPLHPVVPDLPSSVVFPDREAARAVLEECTSFIPEARAVLDLVDQCPKQI 247
         CA V+P  EP LHP++PDL +S VF DR+AARAVLEECTSFIPEARAVLDLVDQCPK++
Sbjct: 136 DCACVLPAQEPHLHPLLPDLLNSAVFQDRDAARAVLEECTSFIPEARAVLDLVDQCPKEV 195

Query: 248 QKGKFQVVAIEGLDATGKTTVTQSVADSLKAVLLKSPPSCIGQWRKIFDDEPTIIRRAFY 307
        QKGKFQV+AIEGLDATGKTT+TQSV++SLKAVLL+SPP CI QWRKIFDDEPTIIRRAFY
Sbjct: 196 QKGKFQVIAIEGLDATGKTTLTQSVSESLKAVLLQSPPPCISQWRKIFDDEPTIIRRAFY 255

Query: 308 SLGNYIVASEIAKESAKSPVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVYQWPEDLL 367
        SLGNY+VASEIAKES   PVIVDRYWHSTATYAIATEVSGGLQ+LPPAHHPVYQWP DLL
Sbjct: 256 SLGNYLVASEIAKESTNFPVIVDRYWHSTATYAIATEVSGGLQYLPPAHHPVYQWPGDLL 315

Query: 368 KPDLILLLTVSPEERLQRLQGRGMEKTREEAELEANSVFRQKVEMSYQRMENPGCHVVDA 427
        KPDL+LLLTV+ EER++RLQGRG EKT+EEAELEAN+VFRQKVEM+YQRMENP CH+VDA
Sbjct: 316 KPDLVLLLTVNSEERVRRLQGRGQEKTKEEAELEANNVFRQKVEMTYQRMENPSCHLVDA 375

Query: 428 SPSREKVLQTVLSLIQNS 445
        SPSRE VLQ VL LIQ+S
Sbjct: 376 SPSRETVLQKVLELIQSS 393


>CRA|224000047885452 /dataset=GENESEQP /org=Homo sapiens /taxon=9606
        /mol_type=aa /date=18-DEC-03 /length=251
        /altid=derwent_ac|ADC39138 /altid=derwent_id|ADC39138
        /def=Novel human NOVX polypeptide  SEQ ID NO: 80
```

8

```
                /patent=WO2003010327-A2 /pat_section=Claim
           Length = 251

  Score =  455 bits (1171), Expect = e-127
  Identities = 229/232 (98%), Positives = 229/232 (98%)

Query: 218  AARAVLEECTSFIPEARAVLDLVDQCPKQIQKGKFQVVAIEGLDATGKTTVTQSVADSLK 277
            A  AVLEECTSFIPEARAVLDLVDQCPKQIQKGKFQVVAIEGLDATGKTTVTQS ADSLK
Sbjct: 20   ALGAVLEECTSFIPEARAVLDLVDQCPKQIQKGKFQVVAIEGLDATGKTTVTQSAADSLK 79

Query: 278  AVLLKSPPSCIGQWRKIFDDEPTIIRRAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTA 337
            AVLLKSPPSCIGQWRKIFDDEPTIIRRAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTA
Sbjct: 80   AVLLKSPPSCIGQWRKIFDDEPTIIRRAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTA 139

Query: 338  TYAIATEVSGGLQHLPPAHHPVYQWPEDLLKPDLILLLTVSPEERLQRLQGRGMEKTREE 397
            TYAIATEVSGGLQHLPPAHHPVYQWPEDLLKPDLILLLTVSPEERLQRLQGRGMEKTREE
Sbjct: 140  TYAIATEVSGGLQHLPPAHHPVYQWPEDLLKPDLILLLTVSPEERLQRLQGRGMEKTREE 199

Query: 398  AELEANSVFRQKVEMSYQRMENPGCHVVDASPSREKVLQTVLSLIQNSFSEP 449
            AELEANSVFRQKVEMSYQRMENPGCHVVDASPSREKVLQTVLSLIQNSFSEP
Sbjct: 200  AELEANSVFRQKVEMSYQRMENPGCHVVDASPSREKVLQTVLSLIQNSFSEP 251


>CRA|223000060714134 /dataset=GENESEQP /org=Homo sapiens /taxon=9606
           /mol_type=aa /date=26-FEB-04 /length=290
           /altid=derwent_ac|ADF76971 /altid=derwent_id|ADF76971
           /def=Novel human secreted and transmembrane protein
           SeqID 646 /patent=WO2003072035-A2 /pat_section=Claim
           Length = 290

  Score =  441 bits (1133), Expect = e-123
  Identities = 229/252 (90%), Positives = 232/252 (92%), Gaps = 4/252 (1%)

Query: 202  PVVPDLPS-SVVFPDREAARAVLEECTSF---IPEARAVLDLVDQCPKQIQKGKFQVVAI 257
            P+   LPS V  +R  R     CT +    IPEARAVLDLVDQCPKQIQKGKFQVVAI
Sbjct: 39   PIWATLPSWRVSKRERAIPRRGKSLCTVYLLYIPEARAVLDLVDQCPKQIQKGKFQVVAI 98

Query: 258  EGLDATGKTTVTQSVADSLKAVLLKSPPSCIGQWRKIFDDEPTIIRRAFYSLGNYIVASE 317
            EGLDATGKTTVTQSVADSLKAVLLKSPPSCIGQWRKIFDDEPTIIRRAFYSLGNYIVASE
Sbjct: 99   EGLDATGKTTVTQSVADSLKAVLLKSPPSCIGQWRKIFDDEPTIIRRAFYSLGNYIVASE 158

Query: 318  IAKESAKSPVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVYQWPEDLLKPDLILLLTV 377
            IAKESAKSPVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVYQWPEDLLKPDLILLLTV
Sbjct: 159  IAKESAKSPVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVYQWPEDLLKPDLILLLTV 218

Query: 378  SPEERLQRLQGRGMEKTREEAELEANSVFRQKVEMSYQRMENPGCHVVDASPSREKVLQT 437
            SPEERLQRLQGRGMEKTREEAELEANSVFRQKVEMSYQRMENPGCHVVDASPSREKVLQT
Sbjct: 219  SPEERLQRLQGRGMEKTREEAELEANSVFRQKVEMSYQRMENPGCHVVDASPSREKVLQT 278

Query: 438  VLSLIQNSFSEP 449
            VLSLIQNSFSEP
Sbjct: 279  VLSLIQNSFSEP 290


>CRA|114000021936825 /dataset=GENESEQP /org=Mus musculus
           /taxon=10090 /mol_type=aa /date=07-MAR-02 /length=431
           /altid=derwent_ac|ABB57297 /altid=derwent_id|ABB57297
           /def=Mouse ischaemic condition related protein sequence
           SEQ ID NO:834 /patent=WO200188188-A2 /pat_section=Claim
           Length = 431

  Score =  259 bits (663), Expect = 2e-68
  Identities = 133/192 (69%), Positives = 152/192 (79%), Gaps = 9/192 (4%)
```

```
Query: 225  ECTSFIPEARAVLDLVDQCPKQIQKGKFQVVAIEGLDATGKTTVTQSVADSLKAVLLKSP 284
            +CTSFIPEARAVLDLVDQCPK++QKGKFQV+AIEGLDATGKTT+TQ         +
Sbjct: 157  QCTSFIPEARAVLDLVDQCPKEVQKGKFQVIAIEGLDATGKTTLTQHFKSLSRLSSYSRH 216


Query: 285  PSCIGQWRKIFDDEPTIIR-----RAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTATY 339
            P CI   K +++  ++       F   NY+VASEIAKES    PVIVDRYWHSTATY
Sbjct: 217  PPCI----KPVEEDLLMMNLLSFEEPFILWANYLVASEIAKESTNFPVIVDRYWHSTATY 272


Query: 340  AIATEVSGGLQHLPPAHHPVYQWPEDLLKPDLILLLTVSPEERLQRLQGRGMEKTREEAE 399
            AIATEVSGGLQ+LPPAHHPVYQWP DLLKPDL+LLLTV+ EER++RLQGRG EKT+EEAE
Sbjct: 273  AIATEVSGGLQYLPPAHHPVYQWPGDLLKPDLVLLLTVNSEERVRRLQGRGQEKTKEEAE 332


Query: 400  LEANSVFRQKVE 411
            LEAN+VFRQKVE
Sbjct: 333  LEANNVFRQKVE 344



>CRA|224000046444908 /dataset=GENESEQP /org=Mus musculus
            /taxon=10090 /mol_type=aa /date=18-DEC-03 /length=508
            /altid=derwent_ac|ADC63372 /altid=derwent_id|ADC63372
            /def=Mouse thymidylate kinase /patent=US2003087294-A1
            /pat_section=Disclosure;
          Length = 508


 Score =  199 bits (506), Expect = 3e-50
 Identities = 96/120 (80%), Positives = 104/120 (86%)


Query: 220  RAVLEECTSFIPEARAVLDLVDQCPKQIQKGKFQVVAIEGLDATGKTTVTQSVADSLKAV 279
            +   ++CTSFIPEARAVLDLVDQCPK++QKGKFQV+AIEGLDATGKTT+TQ        +
Sbjct: 97   KTTTQQCTSFIPEARAVLDLVDQCPKEVQKGKFQVIAIEGLDATGKTTLTQHFKSLSRLS 156


Query: 280  LLKSPPSCIGQWRKIFDDEPTIIRRAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTATY 339
                  PSCIGQWRKIFDDEPTIIRRAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTATY
Sbjct: 157  SYSRHPSCIGQWRKIFDDEPTIIRRAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTATY 216




 Score =  178 bits (452), Expect = 5e-44
 Identities = 88/100 (88%), Positives = 89/100 (89%)


Query: 306  FYSLGNYIVASEIAKESAKSPVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVYQWPED 365
            F   NY+VASEIAKES    PVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVYQWPED
Sbjct: 269  FILWANYLVASEIAKESTNFPVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVYQWPED 328


Query: 366  LLKPDLILLLTVSPEERLQRLQGRGMEKTREEAELEANSV 405
            LLKPDLILLLTVSPEERLQRLQGRGMEKTREEAE  A  V
Sbjct: 329  LLKPDLILLLTVSPEERLQRLQGRGMEKTREEAEAIATEV 368




 Score =  132 bits (333), Expect = 3e-30
 Identities = 63/73 (86%), Positives = 70/73 (95%)


Query: 340  AIATEVSGGLQHLPPAHHPVYQWPEDLLKPDLILLLTVSPEERLQRLQGRGMEKTREEAE 399
            AIATEVSGGLQ+LPPAHHPVYQWP DLLKPDL+LLLTV+ EER++RLQGRG EKT+EEAE
Sbjct: 414  AIATEVSGGLQYLPPAHHPVYQWPGDLLKPDLVLLLTVNSEERVRRLQGRGQEKTKEEAE 473


Query: 400  LEANSVFRQKVEM 412
            LEANSVFRQKVE+
Sbjct: 474  LEANSVFRQKVEL 486
```

```
 Score =  120 bits (300), Expect = 2e-26
 Identities = 64/73 (87%), Positives = 66/73 (90%), Gaps = 1/73 (1%)

Query: 225 ECTSFIPEARAVLDLVDQCPKQIQKGKFQVVAIEGLDATGKTTVTQSVADSLKAVLLKSP 284
           ECTSFIPEARAVLDLVDQCPKQIQKGKFQVVAIEGLDATGKTTVTQSVADSLKAVLLKSP
Sbjct: 1   ECTSFIPEARAVLDLVDQCPKQIQKGKFQVVAIEGLDATGKTTVTQSVADSLKAVLLKSP 60

Query: 285 -PSCIGQWRKIFD 296
            S I + R + D
Sbjct: 61  CTSFIPEARAVLD 73
```

```
   Database: geneseqp_all.fasta
     Posted date:  Feb 27, 2004  4:29 PM
   Number of letters in database: 273,213,384
   Number of sequences in database:  1,546,707

Lambda     K        H
   0.321    0.138     0.426

Gapped
Lambda     K        H
   0.267   0.0410     0.140


Matrix: BLOSUM62
Gap Penalties: Existence: 11, Extension: 1
Number of Hits to DB: 273,243,944
Number of Sequences: 1546707
Number of extensions: 12474211
Number of successful extensions: 32890
Number of sequences better than 1.0e-08: 9
Number of HSP's better than  0.0 without gapping: 9
Number of HSP's successfully gapped in prelim test: 0
Number of HSP's that attempted gapping in prelim test: 32867
Number of HSP's gapped (non-prelim): 17
length of query: 449
length of database: 273,213,384
effective HSP length: 119
effective length of query: 330
effective length of database: 89,155,251
effective search space: 29421232830
effective search space used: 29421232830
T: 11
A: 40
X1: 16 ( 7.4 bits)
X2: 38 (14.6 bits)
X3: 64 (24.7 bits)
S1: 41 (21.8 bits)
S2: 148 (61.6 bits)


NCBI pataa Blast Report:

BLASTP 2.2.5 [Nov-16-2002]
Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer,
Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997),
"Gapped BLAST and PSI-BLAST: a new generation of protein database search
programs", Nucleic Acids Res. 25:3389-3402.

Query= CL001305, SEQ ID NO:2, March 3, 2004
        (449 letters)
```

```
Database: pataa.fasta
          143,807 sequences; 24,539,066 total letters


                                                              Score      E
Sequences producing significant alignments:                  (bits) Value
CRA|224000047562998 /altid=gi|27309063 /def=gb|AAO00504.1| Seque...   914    0.0
CRA|224000047566517 /altid=gi|27309064 /def=gb|AAO00505.1| Seque...   199    3e-51


>CRA|224000047562998 /altid=gi|27309063 /def=gb|AAO00504.1| Sequence
           2 from patent US 6489153 /org= /taxon= /div=
           /dataset=pataa /length=449
          Length = 449

 Score =  914 bits (2362), Expect = 0.0
 Identities = 449/449 (100%), Positives = 449/449 (100%)

Query: 1     MAFARRLLRGPLSGPLLGRRGVCAGAMAPPCRFVLELPDCTLAHFALGADAPGDADAPDP 60
             MAFARRLLRGPLSGPLLGRRGVCAGAMAPPCRFVLELPDCTLAHFALGADAPGDADAPDP
Sbjct: 1     MAFARRLLRGPLSGPLLGRRGVCAGAMAPPCRFVLELPDCTLAHFALGADAPGDADAPDP 60

Query: 61    RLAALLGPPERSYSLCVPVTPDAGCGARVRAARLHQRLLHQLRRGPFQRCQLLRLLCYCP 120
             RLAALLGPPERSYSLCVPVTPDAGCGARVRAARLHQRLLHQLRRGPFQRCQLLRLLCYCP
Sbjct: 61    RLAALLGPPERSYSLCVPVTPDAGCGARVRAARLHQRLLHQLRRGPFQRCQLLRLLCYCP 120

Query: 121   GGQAGGAQQGFLLRDPLDDPDTRQALLELLGACQEAPRPHLGEFEADPRGQLWQRLWEVQ 180
             GGQAGGAQQGFLLRDPLDDPDTRQALLELLGACQEAPRPHLGEFEADPRGQLWQRLWEVQ
Sbjct: 121   GGQAGGAQQGFLLRDPLDDPDTRQALLELLGACQEAPRPHLGEFEADPRGQLWQRLWEVQ 180

Query: 181   DGRRLQVGCAQVVPVPEPPLHPVVPDLPSSVVFPDREAARAVLEECTSFIPEARAVLDLV 240
             DGRRLQVGCAQVVPVPEPPLHPVVPDLPSSVVFPDREAARAVLEECTSFIPEARAVLDLV
Sbjct: 181   DGRRLQVGCAQVVPVPEPPLHPVVPDLPSSVVFPDREAARAVLEECTSFIPEARAVLDLV 240

Query: 241   DQCPKQIQKGKFQVVAIEGLDATGKTTVTQSVADSLKAVLLKSPPSCIGQWRKIFDDEPT 300
             DQCPKQIQKGKFQVVAIEGLDATGKTTVTQSVADSLKAVLLKSPPSCIGQWRKIFDDEPT
Sbjct: 241   DQCPKQIQKGKFQVVAIEGLDATGKTTVTQSVADSLKAVLLKSPPSCIGQWRKIFDDEPT 300

Query: 301   IIRRAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVY 360
             IIRRAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVY
Sbjct: 301   IIRRAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVY 360

Query: 361   QWPEDLLKPDLILLLTVSPEERLQRLQGRGMEKTREEAELEANSVFRQKVEMSYQRMENP 420
             QWPEDLLKPDLILLLTVSPEERLQRLQGRGMEKTREEAELEANSVFRQKVEMSYQRMENP
Sbjct: 361   QWPEDLLKPDLILLLTVSPEERLQRLQGRGMEKTREEAELEANSVFRQKVEMSYQRMENP 420

Query: 421   GCHVVDASPSREKVLQTVLSLIQNSFSEP 449
             GCHVVDASPSREKVLQTVLSLIQNSFSEP
Sbjct: 421   GCHVVDASPSREKVLQTVLSLIQNSFSEP 449



>CRA|224000047566517 /altid=gi|27309064 /def=gb|AAO00505.1| Sequence
           4 from patent US 6489153 /org= /taxon= /div=
           /dataset=pataa /length=508
          Length = 508

 Score =  199 bits (506), Expect = 3e-51
 Identities = 96/120 (80%), Positives = 104/120 (86%)

Query: 220   RAVLEECTSFIPEARAVLDLVDQCPKQIQKGKFQVVAIEGLDATGKTTVTQSVADSLKAV 279
             +    ++CTSFIPEARAVLDLVDQCPK++QKGKFQV+AIEGLDATGKTT+TQ        +
Sbjct: 97    KTTTQQCTSFIPEARAVLDLVDQCPKEVQKGKFQVIAIEGLDATGKTTLTQHFKSLSRLS 156

Query: 280   LLKSPPSCIGQWRKIFDDEPTIIRRAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTATY 339
```

12

```
                    PSCIGQWRKIFDDEPTIIRRAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTATY
Sbjct: 157 SYSRHPSCIGQWRKIFDDEPTIIRRAFYSLGNYIVASEIAKESAKSPVIVDRYWHSTATY 216
```

 Score = 178 bits (452), Expect = 5e-45
 Identities = 88/100 (88%), Positives = 89/100 (89%)

```
Query: 306 FYSLGNYIVASEIAKESAKSPVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVYQWPED 365
           F    NY+VASEIAKES    PVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVYQWPED
Sbjct: 269 FILWANYLVASEIAKESTNFPVIVDRYWHSTATYAIATEVSGGLQHLPPAHHPVYQWPED 328

Query: 366 LLKPDLILLLTVSPEERLQRLQGRGMEKTREEAELEANSV 405
           LLKPDLILLLTVSPEERLQRLQGRGMEKTREEAE  A  V
Sbjct: 329 LLKPDLILLLTVSPEERLQRLQGRGMEKTREEAEAIATEV 368
```

 Score = 132 bits (333), Expect = 3e-31
 Identities = 63/73 (86%), Positives = 70/73 (95%)

```
Query: 340 AIATEVSGGLQHLPPAHHPVYQWPEDLLKPDLILLLTVSPEERLQRLQGRGMEKTREEAE 399
           AIATEVSGGLQ+LPPAHHPVYQWP DLLKPDL+LLLTV+ EER++RLQGRG EKT+EEAE
Sbjct: 414 AIATEVSGGLQYLPPAHHPVYQWPGDLLKPDLVLLLTVNSEERVRRLQGRGQEKTKEEAE 473

Query: 400 LEANSVFRQKVEM 412
           LEANSVFRQKVE+
Sbjct: 474 LEANSVFRQKVEL 486
```

 Score = 120 bits (300), Expect = 2e-27
 Identities = 64/73 (87%), Positives = 66/73 (90%), Gaps = 1/73 (1%)

```
Query: 225 ECTSFIPEARAVLDLVDQCPKQIQKGKFQVVAIEGLDATGKTTVTQSVADSLKAVLLKSP 284
           ECTSFIPEARAVLDLVDQCPKQIQKGKFQVVAIEGLDATGKTTVTQSVADSLKAVLLKSP
Sbjct: 1   ECTSFIPEARAVLDLVDQCPKQIQKGKFQVVAIEGLDATGKTTVTQSVADSLKAVLLKSP 60

Query: 285 -PSCIGQWRKIFD 296
            S I + R + D
Sbjct: 61  CTSFIPEARAVLD 73
```

13

```
Number of HSP's better than   0.0 without gapping: 2
Number of HSP's successfully gapped in prelim test: 0
Number of HSP's that attempted gapping in prelim test: 2873
Number of HSP's gapped (non-prelim): 9
length of query: 449
length of database: 24,539,066
effective HSP length: 103
effective length of query: 346
effective length of database: 9,726,945
effective search space: 3365522970
effective search space used: 3365522970
T: 11
A: 40
X1: 16 ( 7.4 bits)
X2: 38 (14.6 bits)
X3: 64 (24.7 bits)
S1: 41 (21.8 bits)
S2: 140 (58.5 bits)
```